

Understanding the Preservation Challenge of Digital Television

Mary Ide, Dave MacCarn, Thom Shepard, and Leah Weisse
WGBH Educational Foundation

EXECUTIVE SUMMARY

By nature and necessity, public broadcasting is a hodgepodge of media types and formats. A documentary might include moving and still images, speeches and voice-overs, sound effects, or a song. Children's programming might include a combination of live-action, cartoons, musical numbers, and kaleidoscopic effects. Source material for any of these production elements might be analog (a strip of film, a track from a 78 phonograph) or digital (panoramic portraits, credit rolls, or logos).

In whatever manifestations these objects previously existed, they become bits and bytes before they reach the public eye. That is an enormous amount of digital information to manage over time. A single second of uncompressed high definition digital content would take up 150 megabytes of storage space. A minute would fill a home computer's 10-gigabyte hard drive. Although the holding capacity per unit volume doubles almost every two years, these technical advancements come at a cost: media obsolescence.

As we move into the increasingly complex digital world, those charged with preserving our television heritage have the opportunity to develop and establish better coordinated and standardized preservation policies and practices to assure what television programs and related assets survive.

INTRODUCTION: STATEMENT OF PROBLEM

In many respects, the dilemma of archiving digital content is the same as it was for analog: how do we preserve the substance of media while its physical containers decay or grow obsolete? For analog products, standard practice recommends procuring appropriate shelf space within a controlled environment. Digital objects may be handled in similar fashion—that is, as shelved artifacts—but this approach avoids examining the qualities that make digital both attractive and perilous for productions. Alternative digital storage solutions are being marketed all the time. Each new option brings its own set of pitfalls as well as rewards. The bottom line: the storage industry has yet to solve the problem of technical obsolescence with the creation of an archive format.

Standard archival practice continues to advocate the refreshing of physical media. Refreshment strategies, which include migration and emulation, may prove effective for some types of media, but they are inadequate for handling the intricacies, interdependencies, and sheer volume of television content.

Over the past decade, television production and broadcasting have been moving from analog to digital. The analog method, which transmits sounds and pictures through continuous wavelike signals or pulses of varying intensity, is being replaced by digital capture and transmission in which sounds and images are converted into groups of binary code (ones and zeros). This transition is both complex and clouded. Materials collected or generated for a television show may consist of a great threaded mesh of digital and analog components, so tightly bound together that, at any point in their life

cycle, one may serve as a surrogate for another. What is analog today could be digital tomorrow. What is digital today may be stored as analog.

A look at the life cycle of a “production object” reveals myriad routes from the capture of the moving image to the airing of the broadcast. Footage is shot in a studio or on location and makes its way into a video editing system. If the source material is analog, a digital capture card converts the analog information into digital signals. Stills may be scanned from photographs and illustrations, then manipulated with software. What starts as a static image can end up as animation. A slow pan across a Civil War battlefield, a zoom into Mary Lincoln’s eyes ... these become simulated camera movements, and the digital object that began as a JPEG (Joint Photographic Expert Group) or TIFF (Tag Image File Format) becomes an MPEG (Motion Picture Experts Group) video file.

Sound or audio tracks are also treated as distinctive elements in a television production. Whether it is background music, a voiceover, or the sound of water dripping, audio tracks must be maintained both as parts of the completed program and as entities unto themselves. The very same audio information might exist as a WAV file and be packaged within an MPEG.

In addition to materials that have clear analog sources, some materials may be created on desktop machines by teams of artists, designers, and computer programmers using a wide range of off-the-shelf software. A program logo, for example, may begin life as a PhotoShop bitmap. It may then be transformed into an Illustrator vector graphic. This vector graphic may be imported into another application, rendered as a 3D moving object, and incorporated into a show.

The very concept of a “finished program” is debatable. We have already witnessed the rising popularity of DVD feature film “extras”: out-takes, cut segments, director’s cut, and alternative endings. Considering that an audience may see as little as 5% of the original footage shot for any given broadcast, there is an enormous long-term potential market in providing them some leftovers. What remains to be explored is the full value of the original source materials for non-fiction productions: unedited interviews or other documentary footage that lends itself to new interpretations as events unfold. We cannot predict the educational or entertainment value that audiences will derive from production materials, but current trends indicate that there is wisdom in saving it all.

How are Items Selected for Collection and Preservation?

Radio and television broadcasting has been a major influence in shaping the political, social, cultural, and economic trends of the twentieth century. Broadcasting has heightened citizen awareness of our global community and its diversity. The broadcast industry’s recordings and related production materials are primary sources for the study of the century’s history and culture. Media mirrors the world; it changes our perception of the world and draws us into that world. Television “... is not just a new way of doing old things but a radically different way of seeing and interpreting the world” (Kernan 1990, 151).

Current appraisal methodologies used to select television programs for preservation suggest a hybrid of the methods traditionally applied to textual materials. Appraisal for selection requires a significant level of knowledge about the moving image production process and analog and digital production technologies. The appraisal criteria must also take into consideration the technical and financial preservation commitment

implications. The fragility of moving images and the rapid advancements in reformatting technologies complicate the ethical and practical accessioning and appraisal process.

Guidelines or standards for selecting television material for preservation exist and are valuable resources. One of the earliest and most comprehensive international television appraisal studies was the 1983 Record and Archives Management Programme (RAMP) study, prepared for the United Nations Educational, Scientific, and Cultural Organization (UNESCO) by Sam Kula. In his RAMP report, Kula acknowledged that selection criteria tend to first meet the needs of broadcasters, and the potential for re-use of programming content is particularly important. Re-use potential also considers the intrinsic historical or cultural value of content (Kula 1990).

The Federation Internationale des Archives de Television/International Federation of Television Archives (FIAT/IFTA) is a European-based international organization of archivists who manage television archival material. FIAT developed the following criteria for master television program selection in 1996:

- material of historic interest in all fields
- material as a record of a place, an object, or national phenomenon
- interview material of historic importance
- interview material indicative of opinions or attitudes of the time
- fictional and entertainment material of artistic interest
- fictional and entertainment material illustrative of social history
- any material, including commercial and presentational, illustrative of the development of television practices and techniques (LC Study 1997, 189).

Commercial and public broadcasting stations and other collecting institutions have developed their selection criteria based on their institutional needs and missions. But for any collecting institution, the preservation commitment is staggering in cost and maintenance—whether it is digital or analog. The time has come to encourage and explore the concept of regional and national planning for the preservation of broadcast television programming.

The Library of Congress study, *Television and Video Preservation 1997: A Study of the Current State of American Television and Video Preservation*, outlines the state of American preservation practices and calls for a concerted national and regional effort to plan for the preservation of American television programming. Librarian of Congress James H. Billington says in the study's preface that, "... at present, chance determines what television programs survive. Future scholars will have to [rely] on incomplete evidence when they assess the achievements and failures of our culture" (LC Study 1997, xi).

STANDARD FORMATS FOR DIGITAL TELEVISION

Standards for digital television include not only the formats for the physical media but also for the broadcast stream itself. The current analog broadcast standard, for example, has an image resolution of 525 horizontal lines and 640 vertical lines or pixels. To understand what this means, consider that your home computer's monitor is likely to have a resolution of 800 by 600 or better. In contrast, the standard resolution for High Definition Television (HDTV) is 1080 lines and 1920 pixels. In addition, the aspect ratio for HDTV is 16:9, while the standard for conventional TV is 4:3. As the numbers suggest, HDTV holds a great deal of promise for today's viewing audience, yet increases the

amount of information available. These numbers also point to a problem: How do you transport this extra information through the same broadcast pipeline?

The ATSC Digital Television Standard (A-53) was devised to increase the amount of broadcast information allowable through a conventional 6 MHz channel. A finished program might be transported directly from an editing station, set up in the control room as a compressed MPEG-2 video file, broadcast to home analog television sets, and may additionally be transferred to some archival storage system or media. Although the A-53 standard is regulated across the United States, the problems of physical storage for this material grow increasingly more complex.

Since 1987, at least 17 digital videotape formats have come into the marketplace, and, as with analog tape, competing and non-compatible formats proliferate. The format issue alone is a nightmare for collecting institutions for two reasons: formats are platform-dependent to particular playback machines, and physical media require constant migration to new formats.

Videotape is a notoriously fragile medium made up of three major components: the backing, the magnetic coating, and the binder that holds the magnetic coating to the backing. While the life expectancy of videotape is, at best, 15-20 years, time and experience have shown that the older analog videotape formats are sturdier and last longer than newer ones.

Some digital video formats use compression. Compression can dramatically reduce the size of a data file by eliminating redundant information by taking advantage of the psycho-visual studies of human perception. Some compression techniques are proprietary. Because manufacturer's implementations vary, they produce "... unanticipated consequences such as a phenomenon called 'concatonation,' in which artifacts of the compression process make it difficult to transfer content to new formats" (Liroff 2001, 8).

While the specifications for DVD were being hammered out, hopes were high in the archival community that it might serve as an adequate preservation vehicle. Now, the consensus among moving image archivists is more pessimistic. Though regarded as an advancement in distribution and access, the DVD disc, like the CD and the CD-ROM disc that it physically resembles, is also subject to deterioration from oxidation, humidity, and physical damage. In addition, there is no guarantee that the format will not become obsolete within another generation. That said, technologies and materials might improve to the extent that the DVD format might be re-evaluated by the archival community. Perhaps a "backward compatible" DVD format might be developed for purely archival use.

ORGANIZATIONAL ISSUES

Organizational issues concerning digital television content include asset and rights management, distribution channels, and user purposes and needs. Solutions to these issues will vary with an institution's mission. Because we are in a transition period of analog to digital, traditional and non-traditional methods of dealing with organizational issues are currently used in tandem.

Asset and Rights Management

Over the past twenty years, an expanding market for production repurposing has encouraged the practice of keeping edited master programs and related production elements. Also, the advent of smaller tape formats has allowed us to store more individual items. Digital Asset Management (DAM) systems provide access and storage to these rich media assets, which are digitally indexed and often associated to specific rights management information.

Digital Rights Management (DRM) entails tracking rights of each creating entity, controlling access, security issues, collecting payments, and distribution. A producing entity must track copyright-related data including insurance agreements, trademark issues, talent payments, licensing and market agreements, co-production payments, and financial support.

The breakdown of program material into segments is crucial to rights management. Segmentation is not only vertical but also horizontal. Attributes must be logged for each component part. For example, music or narration for a program needs to be available as a standalone component, if only to allow editors to remove it for rebroadcast. Rights information needs to be applied to each of these components.

Product placement through digital manipulation may factor into how we manage moving image materials. Though highly controversial, experiments are under way in commercial television to set up product placement variables within dramatic scenes. Flexibility in product placement may be particularly lucrative when a show is licensed for syndication. For example, one version might show a can of Pepsi as a strategically placed prop. In another market, that soda might be digitally turned into Coke. Though it is hard to imagine the public affected by product placement, it is conceivable that just as cable markets license our programs, we may indeed see product placement as a requirement for licensing.

Distribution

There are multiple program distribution routes including broadcast transmission, home video, satellite, cable, and Webcasting. By the year 2003, the Federal Communications Commission has mandated that all commercial and public broadcasting stations will have to convert to the digital television (DTV) transmission standard. Once digital TV is widespread, broadcast materials will exist in several versions and formats. DTV will expand broadcasting capabilities to include three formats: high definition television (HDTV), multicasting, and datacasting. The highest quality will be high definition (HDTV), providing an image far superior to what can be received on analog sets.

Multicasting would permit multiple programs to be carried by one broadcast signal, allowing the broadcasters (like cable systems) to increase the amount of programming available as well as target viewer demographics. It could also allow viewers to experience alternative angles of a particular broadcast. Live drama, breaking news events, and sports telecasts would benefit from multicasting.

Datacasting, as its name implies, allows data (video, audio, text, graphics, maps, and services) to be embedded in the broadcast signal for downloading into a computer or set-top box, allowing the broadcasting of ancillary materials to accompany a program. These materials may be accessible in a number of ways: as downloadable data that may be collected and accessed through computers, or as streaming content that may be

viewed on a predesignated portion of a television screen. Datacasting could give viewers immediate access to a wealth of supplementary material, such as cast lists, biographies, and transcripts. These features are like the “extras” that are included in many current DVDs.

New technologies continue to up the ante for audience expectations. Today, we want our video on demand. Tomorrow, we will have a side order of metadata. As long as there are audiences hungry for both quantities and varieties of information, there will be industries to supply those needs. As television grows more Web-like, providing quick and easy access to lots of digital information through digital hyperlinks, those charged with the preservation and access to television content will play a key role, and perhaps in the process will finally win public recognition for their efforts.

Users

A measure of how the public uses digital assets is reflected in the coined term, “edutainment.” The expression has caught on throughout the world and is used in several languages. Literally, it is the melding of the words “education” and “entertainment.” Figuratively, it means “learning that is fun.” What is often missing in academic discussions of electronic information is the “fun factor.” Even tools for data retrieval, for example, are not only getting more attractive but also easier for the general public to use.

The user base stretches beyond the general public; education professionals, researchers, the production community, and others have also embraced new technologies. All are benefiting from the use of television production assets created specifically for curriculum research, distance learning, and classroom reference. Moving image collections have been developing Web sites for use by educators such as the WGBH New Television Workshop Project (<http://main.wgbh.org/wgbh/NTW>).

WGBH’s National Center for Accessible Media (NCAM) makes public media accessible to disabled persons, minority language users, people with low literacy skills, and other under-served populations. For example, it offers closed captioning and Descriptive Video Services (DVS) for those with special hearing and sight needs. NCAM researches and develops media access technologies and explores how existing technologies may benefit other populations. These access technologies create another set of production assets.

IMPLICATONS FOR LONG-TERM PRESERVATION

Storage

A clear distinction must be made between how we preserve broadcast materials and how we access them over time. Preserving data is crucial, but how readily available will these materials need to be? Off-line storage takes the longest time to retrieve. It is usually boxed and stored on a shelf but is cataloged and available. Near-line storage provides intermediate access. Near-line storage is linked to the concept of the “jukebox” system, a collection of optical or tape drives that reside in a hardware device consisting of numerous slots or “bays” and a robotic arm. The stored data are not instantly accessed, but instead are retrieved through various human or mechanical means. Online storage provides the most immediate access, typically spinning disk, possibly SAN (Storage Area Network) or NAS (Network Attached Storage), accessible through file

systems and Internet/LANs (Local Area Networks). In hardware terms, online storage is any storage device that is perpetually available to authorized users. Digital storage will be so cheap in years to come that we may easily afford to keep exact copies of our materials in several distinct locations. This “redundant storage” would help protect our assets in times of disaster. On the other hand, limitless storage introduces new problems of access and management.

There are basically two approaches to storing digital video images. We can store whole programs and create databases that contain metadata. And we can store all of the clips that are included in the program as separate files and then rely on Edit Decision Lists (EDLs) to serve as blueprints for our broadcasts. Both options rely on some form of stratification of the media. Stratification is a system of video annotation that uses time-codes to identify marking points within an audio or video object. Descriptions can be linked to these points by storing them together with the time-code information. In the same way that video can contain many tracks, metadata may also have several layers, each with its own set of referenced time-codes. For example, a transcript may occupy one metadata layer, while captioning information may occupy another. Other layers may include DVS material, copyright, or image content description.

Even as storage space becomes limitless and more reliable, we still need to grapple with the problem of software obsolescence. Storing the same information in many different standard and proprietary formats may be one way to protect our assets, but this approach will require a great dependency on software tools to keep track of them. Broadcast materials are built upon a hierarchy: series, program, segment, clip, and even a single frame. Tools will have to be robust enough to manage these materials on all levels. As Howard Besser writes, those concerned with preservation need “...to move away from an artifact-based approach [to preservation] and instead adopt an approach that focuses on stewardship of disembodied digital information” (Besser 2001, 4).

Proposed Solutions

In the archival communities, the debate over digital preservation has focused on three strategies: migration, emulation, and bundling. Migration is the process of moving data from a digital format that is determined to be obsolete to a platform that is currently in use. As a preservation strategy, migration is prone to bad judgment calls. As a technical solution, migration may damage the essence of the material by dropping crucial data that could result in its loss of function or in its original look and feel.

Emulation approaches the problem through a kind of a virtual time machine. To sustain a digital object's original look and feel, it is necessary to emulate the application that created the object, emulate the operating system upon which the application ran, and emulate the hardware platform upon which the operating system was housed. This is not a one-time, fix-all strategy. Emulation software will have its own hardware and operating system dependencies. The virtual time machine itself may have to be emulated.

A problem with emulation specific to audio and image content is the possibility that the original playback application is limited as compared to later versions or other applications. In other words, the application that created the data file may not be the best application for playing it back. A digital media file often contains more information than may be displayed through its current application. For example, a moving image file may be exported from a software application at a greater resolution than the application itself can display. Metadata fields may be hidden from the current application but

available or reserved for future versions. In other words, the emulation time machine may need to know which version of an application most completely captures or extracts the data.

Bundling is the process of bonding metadata with content within the same file format. This bundling may include information about the provenance of a particular item. The Universal Preservation Format (UPF), which was proposed by WGBH, uses a data file mechanism that bundles metadata with the data representing the actual image, sound, or text. The metadata identifies this data "essence" within a registry of standard data types and serves as the source code for mapping or translating binary composition into accessible or useable forms. The UPF is designed to be independent of the computer applications used to create content, independent of the operating system from which these applications originated, and independent of the physical media upon which that content is stored. The UPF is characterized as "self-described" because it includes, within its metadata, all the technical specifications required to build and rebuild appropriate media browsers to access contained materials throughout time.

Other initiatives that use bundling or packaging include the Open Archival Information System (OAIS) and the Digital Rosetta Stone Model.

Longevity Problems

Howard Besser outlines five longevity problems specific to preserving all digital records including video assets. These are: the viewing problem, the translation problem, the custodial problem, the scrambling problem, and the inter-relational problem (Besser, 2000, 156).

The viewing problem is the fact that electronic content is stored on physical devices that deteriorate and require proactive planning to migrate and assure longevity.

The translation problem focuses on understanding that "...work translated into new delivery devices changes meaning" (Besser 2001, 3). A simple example is a motion picture resized for the television screen.

The custodial problem concerns determining who will be responsible for the long term preservation and authentication of digital content; will it be archivists, computer technologists, others or a collaboration of many?

The scrambling problem for digital television is twofold and relates to the compromise of using compression techniques to satisfy limited storage and bandwidth transmission capabilities and encryption schemes to protect content, which make future access potentially a problem. Compression compromises the integrity of original content and encryption adds an additional layer of complexity to a fragile digital object.

The inter-relational problem concerns the complexity of related information to and within a digital object. Because boundaries of information sets or digital objects are not usually defined, this raises not only custodial concerns but also intellectual property concerns.

UNRESOLVED ISSUES

Paul Messier has suggested that an adequate digital video preservation plan should:

1. make a format accessible on standard equipment at various levels of access;
2. capture image at the highest possible quality resolution rate using minimum or no compression;
3. develop guidelines for digital conversion based on the type of source material;
4. use formats and equipment that meet national, and international standards; and
5. ensure a data migration path that is a hedge against format and machine obsolescence (Messier 1996, 3).

Standards for cataloging moving image materials are continually in evolution. The Library of Congress has set the most prevalent standard. Techniques for creating access to digital content on an international scale include the Dublin Core (DC) initiative and MPEG-7, to name a few. The Dublin Core, being developed by international cross-disciplinary groups, is a set of 15-plus basic information metadata fields for identifying content and access points. Working groups within the Dublin Core metadata initiative are proposing enhancements to this basic set of tags that address cataloging needs of specific industries or domains. Currently, these "application profiles" are being proposed for education, libraries, and bibliographic citations, among others. Some researchers have begun to lay the foundation for an application profile for static and moving image and audio files. MPEG-7 is the Multimedia Content Description Interface standard developed by the MPEG, whose goal is to provide a rich set of standardized metadata fields to describe multimedia content.

Ethical issues concern maintaining the integrity of original content and intent; this is particularly acute with digital morphing capabilities to change and manipulate images in ways that cannot be detected. Included in this dilemma is compression of files that can compromise original intent and artistic authenticity. For example, when moving image materials are available only as low-resolution digital files or scanned from older analog formats, pixels might be filled in to give the illusion of a higher density resolution. Finally, there are the issues of adherence to copyright law, protection of privacy rights, and confidentiality.

In the not-too-distant future, the line between moving image distribution and moving image projection may fade completely. Already there have been experiments in which a motion picture was transmitted from a remote location and projected into a movie theater. The first such test occurred on June 6, 2000, when Cisco Systems Inc. joined with Twentieth Century Fox to digitally transmit "Titan A. E." from Burbank, California to the Woodruff Arts Center in Atlanta, Georgia. The notion of an "artifact-free" method of distribution will have a great impact on preservation. Instead of moving digital information to tapes for distribution, data will simply consist of a file transfer to some temporary storage device, which might periodically be wiped clean. Assumptions about who is responsible for preserving these broadcast materials may result in some tremendous losses.

The issue of who is responsible for the preservation of digital content has not been satisfactorily resolved. Preservation of digital content must be a collaborative effort between the professional archivist, technology expert, user, and the creating and producing entity.

Inaction on the preservation front will assure the continued loss of the nation's television heritage. As stated in the Library of Congress study, "all organizations having custody of American television and video materials, whether private or public bodies, should recognize their responsibilities for preserving a part of the historical and cultural heritage" (Study 1, 123).

REFERENCES

- Besser, Howard. 2000. Digital Longevity. In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, edited by Maxine Sitts. Andover, Mass.: Northeast Document Conservation Center.
- Besser, Howard. 2001. Digital Preservation of Moving Image Material? Available from www.gseis.ucla.edu/~howard/Papers/amia-longevity.html.
- Council on Library and Information Resources. 2000. *Authenticity in a Digital Environment*. Washington, D.C.: Council on Library and Information Resources.
- Gilliland-Swetland, Anne J., and Philip B. Eppard. 2000. Preserving the Authenticity of Contingent Digital Objects. In *D-Lib Magazine* 6(7-8). Available from www.dlib.org/dlib/july00/eppard/07eppard.html.
- Gilliland-Swetland, Anne J. 1999. The Long-Term Preservation of Authentic Electronic Records: InterPARES. Speech given on August 28 at Society of American Archivists Annual Meeting in Pittsburgh, Pa.
- Granger, Stewart. 2000. Emulation as a Digital Preservation Strategy. *D-Lib Magazine* 6(10). Available from www.dlib.org/dlib/october00/granger/10granger.html.
- Hunter, Gregory S. 2000. *Preserving Digital Information: A How-to-do-It Manual #93*. New York: Neal-Schuman Publishers, Inc.
- Hunter, Jane. 1999. MPEG-7 Behind the Scenes. *D-Lib Magazine* 5(9). Available from www.dlib.org/dlib/september99/hunter/09hunter.html.
- Kernan, Alvin. 1990. *Death of Literature*. New Haven: Yale University Press.
- Kula, Sam. 1990. Selected Guidelines for the Management of Records and Archives: A RAMP reader. PGI-90/WS/6. Paris: UNESCO. Available from <http://www.unesco.org/webworld/ramp/html/r9006e/r9006e00.htm#Contents>
- Library of Congress. 1997. *Television and Video Preservation: A Report of the Current State of American Television and Video Preservation*. 3 vols. Washington, D.C.: Library of Congress.
- Lindner, Jim. 1998. Digitization Reconsidered. Available from www.vidipax.com/articles/digirecon.html.
- Liroff, David. 2001. Media Asset Management—The Long Term View. Speech given on August 21 at the Sun Microsystems Digital Media Universe, Beverly Hills, Calif.
- MacCarn, Dave. 2000. *Toward A Universal Data Format For the Preservation of Media*. Available from http://info.wgbh.org/upf/papers/SMPTE_UPF_paper.html.
- Messier, Paul. 1996. Criteria for Assessing Digital Video as a Preservation Medium. *Bay Area Video Coalition (BVAC) Playback 1996 [Conference] Report to the Field*.
- National Research Council. 2001. *LC21: A Digital Strategy for the Library of Congress: Executive Summary*. Available from <http://stills.nap.edu/books/0309071445/html/>.

OCLC/RLG Working Group on Preservation Metadata. 2001. Preservation Metadata for Digital Objects: A Review of the State of the Art (January 31).

Sadashige, Koichi, 2000. Data Storage Technology Assessment 2000. Available from <http://www.nta.org/Bibliography/techreports/part1.htm>.

Su-Shing Chen. 2001. The Paradox of Digital Preservation. *Computer* 34(3): 24-28. Available from Institute of Electrical and Electronics Engineers, Inc., #0018-9162/01.

Wheeler, Jim. Video Q&A. *Newsletter of the Association of Moving Image Archivists*. 49;34.